Confidence Intervals in Public Health
Office of Public Health Assessment
Utah Department of Health
1/15/2002
Brian Paoli, Lois Haggard, Gulzar Shah

When public health practitioners use health statistics, sometimes they are interested in the actual number of health events, but more often they use the statistics to assess the *true underlying risk* of a health problem in the community. Observed health statistics, that is, those counts, rates or percentages that are computed or estimated from health surveys, vital statistics registries, or other health surveillance systems, are not always an accurate reflection of the true underlying risk in the population. Observed rates can vary from sample to sample or year to year, even when the true underlying risk remains the same.

Statistics based on samples of a population are subject to *sampling error*. Sampling error refers to random variation that occurs because only a subset of the entire population is sampled and used to estimate a finding for the entire population. It is often mis-termed "margin of error" in popular use. Even those statistics based on health events in an entire population are based on an arbitrary sample of time (e.g., January 1 through December 31) and are thus subject to a certain amount of sampling error. In general, sampling error gets larger when the sample, population or number of events is small.

Statistical sampling theory is used to compute a *confidence interval* to provide an estimate of the potential discrepancy between the true and observed rates. Understanding the potential size of that discrepancy can provide information about how to interpret the observed statistic. For instance, if the state infant death rate of 5.94 increased to 6.03 in a one-year period, is that increase something that should cause concern? If the smoking rate among teens decreased from 13% to 8%, is that cause for celebration?

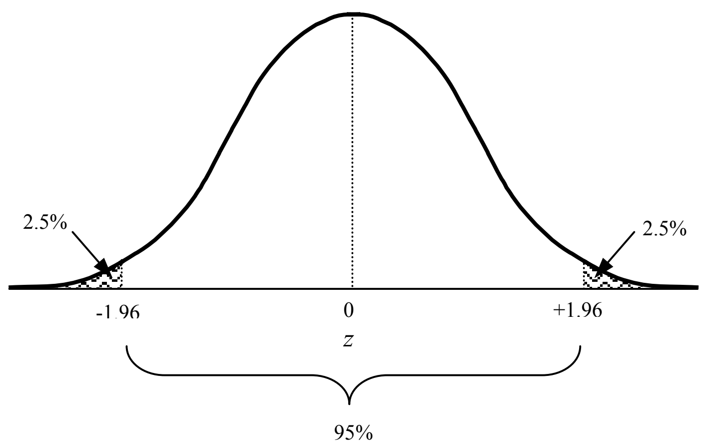
Technically speaking, the 95% confidence interval indicates the range of values within which the statistic would fall 95% of the time if the researcher were to calculate the statistic (e.g., a percentage or rate) from an infinite number of samples of the same size, drawn from the same population. In less technical language, the confidence interval is a range of values within which the “true” value of the rate is expected to occur (with 95% probability). This document describes the most common methods for calculation of 95% confidence intervals for some rates and estimates commonly used in public health.

95% Confidence Interval for a Percentage From a Survey Sample:

Although statistics from survey samples are not necessarily the most commonly-used statistics in public health, they are a good starting point to begin talking about sampling theory. To calculate a confidence interval for a percentage, one must first calculate the *standard error* of the percentage. A percentage is also known as the mean of a binomial distribution. Just in case you were wondering, the standard error of the mean is a measure of dispersion for the hypothetical distribution of means called the *sampling distribution of the mean*. This is a distribution of means calculated from an infinite number of samples of a certain size (the same size as the sample from which the original percentage was measured), drawn from the same population as the original sample.

Once you have calculated the standard error of the percentage, you must decide how large you want the confidence interval to be. The most common alternative is a 95% confidence interval. This is the width of the interval that includes the mean (the *sampling distribution of the mean*, mentioned above) 95% of the time. In a little plainer language, a 95% confidence interval for a percentage is the range of scores within which the percentage will be found if you went back and got a different sample from the same population. At least 95% of the time it will.

Transforming the standard error into a 95% confidence interval is rather simple. Fortunately, the *sampling distribution of the mean* has a shape that is almost identical to what is known as the *normal distribution*.[1] You need only multiply the standard error by the Z-score of the points in the normal distribution that exclude 2.5% of the distribution on either end (two-tailed). That Z-score is 1.96. A Z-score of 1.96 defines the 95% confidence interval. A Z-score of 1.65 defines a 90% confidence interval.



standard error = square root (((rate*(1-rate)) /sample n))

Example: 13% of surveyed respondents indicated that they smoked cigarettes. The sample consisted of 500 persons.

standard error = square root (((.13*(.87)) /500)) = .015

[1] A *distribution* is a tool that is used in statistics to associate a statistic (e.g., a percentage, average, or other statistic) with its probability. When researchers talk about a measure being "statistically significant," they have used a *distribution* to evaluate the probability of the statistic, and found that it would be improbable under ordinary conditions. In most cases, we can rely on measures such as rates, averages, and proportions as having an underlying *normal distribution*, at least when the sample size is large enough.

If you have another measure, such as an average, you must modify the formula to get the correct answer.

Then the 95% confidence interval is:

the percentage $\pm$ 1.96 * standard error = .13 -/+ 1.96*.015 = .0294, or 2.94%
so the 95% confidence interval is 10.06% to 15.94%

Small Samples

If the sample from which the percentage was calculated was rather small, say, smaller than 60, then the shape of the *sampling distribution of the mean* is not the same as the shape of the *normal distribution*. In this special case, we can use another distribution, known as the *t* distribution, that has a slightly different shape than the *normal distribution*.

The procedures in this case are analogous to those above but the *t*-score comes from a family of distributions which depend on the "degrees of freedom." The number of degrees of freedom is defined as "n-1" where "n" is the size of the sample. For a sample of size=30 the degrees of freedom is equal to 29. So, for a 95% confidence interval, you must use the *t* -score associated with 29 degrees of freedom. That particular *t*-score is 2.045 (see Table 1.). So you would multiply the standard error by 2.045 instead of 1.96 to generate the 95% confidence interval.

If our sample were a different size, say 20, then the degrees of freedom would be 19, which is associated with a *t*-score of 2.093 for a 95% confidence interval. As you see the interval will get wider as our sample size is reduced. This reflects the uncertainty in our estimate of the variance in the population. For a 95% confidence interval with 9 degrees of freedom the *t*-score is 2.262. Table 1. lists the *t*-scores for specific degrees of freedom and sizes of confidence interval. For a 95% confidence interval, you would use the *t*-score that defines the points on the distribution that excludes the most extreme 5% of the distribution, which is 0.025 on either end of the curve.

Finite Populations

If the survey sampled all or most of the members of the population, then using the finite population correction factor will improve (decrease) the calculated standard error of the mean.

finite population correction factor = 1-f, where f is the sampling fraction
f = n/N, or, simply the percentage of the population that was included in the sample

standard error of the mean for a binomial distribution for a finite sample
= square root (((rate*(1-rate)) /sample n)*(1-f))

When the Percentage is Close to 0% or 100%

When the percentage is close to 0% or 100%, the formulas given above can result in illogical results - confidence limits that fall below 0% or above 100%. A special formula is used to calculate asymmetric confidence limits in these cases.

Complex Sample Designs

The above formulas assume that the survey sample was a simple random sample. If the survey used a complex sample design (such as clustering within households or disproportionate sampling from various geographic regions), special techniques must be used to calculate the standard error of the mean. Those techniques are accomplished using statistical software such as SAS or SUDAAN.

95% Confidence Intervals When the Event Is Rare:

In the case of rare events, the *normal distribution* no longer applies. A different distribution, the *Poisson distribution* is used to model rare events, such as the "100 year flood." It is also used to gauge the probability of infant mortality or cancer. This distribution is not symmetric about its mean and so the associated confidence intervals will not be symmetric (the upper limit is farther from the estimate than is the lower limit, or the "plus" of the ± is larger than the "minus").

The *Poisson distribution* does, however, assume the shape of a *normal distribution* when the number of events is greater than about 100. So we use a *Poisson distribution* for rare events (when the number of events is less than 100), but we can use the normal distribution when the number of events is greater than 100.

In Table 2 you will find lower and upper *confidence factors* for use in calculating a 95% confidence interval for a rate based on a specified number of events, from 1 to 100. To calculate the confidence interval multiply the estimated rate by the *confidence factor* associated with the number of events on which the rate is based.

For example, in a given geographic area, there were 722 births in a single year, and seven infant deaths. The infant mortality rate in was 9.7 per 1,000 live births, calculated as [(7/722)*1,000]. The lower and upper confidence limits are calculated using the *confidence factors* found on Table 2. The factors for seven events are .4021 and 2.0604 for the lower and upper limits of the confidence interval, respectively. The lower limit of the confidence interval = 9.7*.4021 = 3.90, and the upper limit = 9.7*2.0604 = 19.99, for a rate of 9.7 and a 95% confidence interval from 3.90 to 19.99. If this same rate had been based on 100 deaths then the *confidence factors* would be .8136 and 1.2163. The lower limit would be 9.7*.8136, and the upper limit 9.7*1.2163 for an estimate of 9.7 with a confidence interval from 7.89 to 11.80. This interval is much smaller due to the greater number of deaths on which the rate is based.

Age-Adjusted Rates

When comparing across geographic areas, some method of age-adjusting is typically used to control for area-to-area differences in health events that can be explained by differing ages of the area populations. For example, an area that has an older population will have higher crude (not age-adjusted) rates for cancer, even though its exposure levels and cancer rates for specific age groups are the same as those of other areas. One might incorrectly attribute the high cancer rates to some characteristic of the area other than age. Age-adjusted rates control for age effects, allowing better comparability of rates across areas. Direct standardization adjusts the age-specific rates observed in the small area to the age distribution of a standard population (Lilienfeld & Stolley, 1994).

The confidence interval for directly standardized rates (DSR) can be computed as follows:

$$
\begin{aligned}
CI(DSR) &= \pm 1.96 * SE(DSR) * K \\
&= \pm 1.96 * SQRT(VAR(DSR)) * K \\
&= \pm 1.96 * SQRT(\Sigma W_i^2 * Var(R_i)) * K \\
&= \pm 1.96 * SQRT(\Sigma W_i^2 * ((R_i * (1 - R_i))/P_i)) * K
\end{aligned}
$$

Where...

$SE(DSR)$ = the standard error of the directly standardized rate

K = a constant (e.g., 100,000) that is being used to communicate the rate

W_{si}^2 = the population weight for the ith age group in the standard population

R_i = the age-specific death/disease rate in the ith age group of the small area population (# deaths/population count)

P_i = the population count in age group i of the small area

Indirectly Age-Adjusted Rates

The direct method can present problems when population sizes are particularly small. Calculating directly standardized rates requires calculating age-group-specific rates, and for small areas these age-specific rates may be based on one or two events. In such cases, indirect standardization of rates may be used.

Indirectly standardized rates are based on the standard mortality or morbidity ratio (SMR) and the crude rate for a standard population. Indirect standardization adjusts the overall standard population rate to the age distribution of the small area (Lilienfeld & Stolley, 1994). It is technically appropriate to compare indirectly standardized rates only with the rate in the standard population, not with each other.

An indirectly standardized death or disease rate (ISR) can be computed as:

$$ISR = SMR * R_s$$

$$SMR = \frac{\text{observed deaths/disease in the small area}}{\text{expected deaths/disease in the small area}} = \frac{D}{e} = \frac{D}{\Sigma(R_{si} * P_i)}$$

Where...

R_s = the crude death/disease rate in the standard population
R_{si} = the age-specific death/disease rate in age group i of the standard population (# deaths/population count)
P_i = the population count in age group i of the small area

For indirectly standardized rates based on events that follow a Poisson distribution and for which the ratio of events to total population is small (<.3) and the sample size is large, the following two methods can be used to calculate confidence interval (Kahn & Sempos, 1989).

(1) When the number of events >20:

$$CI(ISR) = (SMR \underline{+} 1.96\ SQRT(SMR/e)) * R_s * K$$

Where...

R_s = the crude death/disease rate in the standard population
K = a constant (e.g., 100,000) that is being used to communicate the rate
SMR = $\frac{\text{observed deaths/disease in the small area}}{\text{expected deaths/disease in the small area}}$
e = expected deaths/disease in the small area = $3(R_{si} * P_i)$
R_i = the age-specific death/disease rate in the ith age group of the small area population (# deaths/population count)
P_i = the population count in age group i of the small area

(2) When the number of events <=20:

$$LL(ISR) = (\text{Lower limit for parameter estimate from Poisson table/e})) * R_s * K$$
$$UL(ISR) = (\text{Upper limit for parameter estimate from Poisson table/e})) * R_s * K$$

Where LL is the lower confidence interval limit, and UL is the upper confidence interval limit.

Table 1. Upper critical values of Student's t distribution with < degrees of freedom

Probability of exceeding the critical value

<	0.10	0.05	0.025	0.01	0.001
1	3.078	6.314	12.706	31.821	318.313
2	1.886	2.920	4.303	6.965	22.327
3	1.638	2.353	3.182	4.541	10.215
4	1.533	2.132	2.776	3.747	7.173
5	1.476	2.015	2.571	3.365	5.893
6	1.440	1.943	2.447	3.143	5.208
7	1.415	1.895	2.365	2.998	4.782
8	1.397	1.860	2.306	2.896	4.499
9	1.383	1.833	2.262	2.821	4.296
10	1.372	1.812	2.228	2.764	4.143
11	1.363	1.796	2.201	2.718	4.024
12	1.356	1.782	2.179	2.681	3.929
13	1.350	1.771	2.160	2.650	3.852
14	1.345	1.761	2.145	2.624	3.787
15	1.341	1.753	2.131	2.602	3.733
16	1.337	1.746	2.120	2.583	3.686
17	1.333	1.740	2.110	2.567	3.646
18	1.330	1.734	2.101	2.552	3.610
19	1.328	1.729	2.093	2.539	3.579
20	1.325	1.725	2.086	2.528	3.552
21	1.323	1.721	2.080	2.518	3.527
22	1.321	1.717	2.074	2.508	3.505
23	1.319	1.714	2.069	2.500	3.485
24	1.318	1.711	2.064	2.492	3.467
25	1.316	1.708	2.060	2.485	3.450
26	1.315	1.706	2.056	2.479	3.435
27	1.314	1.703	2.052	2.473	3.421
28	1.313	1.701	2.048	2.467	3.408
29	1.311	1.699	2.045	2.462	3.396
30	1.310	1.697	2.042	2.457	3.385
31	1.309	1.696	2.040	2.453	3.375
32	1.309	1.694	2.037	2.449	3.365
33	1.308	1.692	2.035	2.445	3.356
34	1.307	1.691	2.032	2.441	3.348
35	1.306	1.690	2.030	2.438	3.340
36	1.306	1.688	2.028	2.434	3.333
37	1.305	1.687	2.026	2.431	3.326
38	1.304	1.686	2.024	2.429	3.319
39	1.304	1.685	2.023	2.426	3.313
40	1.303	1.684	2.021	2.423	3.307
41	1.303	1.683	2.020	2.421	3.301
42	1.302	1.682	2.018	2.418	3.296
43	1.302	1.681	2.017	2.416	3.291
44	1.301	1.680	2.015	2.414	3.286
45	1.301	1.679	2.014	2.412	3.281
46	1.300	1.679	2.013	2.410	3.277
47	1.300	1.678	2.012	2.408	3.273
48	1.299	1.677	2.011	2.407	3.269
49	1.299	1.677	2.010	2.405	3.265
50	1.299	1.676	2.009	2.403	3.261
51	1.298	1.675	2.008	2.402	3.258
52	1.298	1.675	2.007	2.400	3.255
53	1.298	1.674	2.006	2.399	3.251
54	1.297	1.674	2.005	2.397	3.248
55	1.297	1.673	2.004	2.396	3.245
56	1.297	1.673	2.003	2.395	3.242
57	1.297	1.672	2.002	2.394	3.239
58	1.296	1.672	2.002	2.392	3.237
59	1.296	1.671	2.001	2.391	3.234
60	1.296	1.671	2.000	2.390	3.232

Table 2. 95% Confidence Interval Factors for Poisson-Distributed Events

number of events	95% Confidence Interval, Lower Limit Factor	95% Confidence Interval, Upper Limit Factor
0	0.0000	3.7000
1	0.0253	5.5716
2	0.1211	3.6123
3	0.2062	2.9224
4	0.2725	2.5604
5	0.3247	2.3337
6	0.3670	2.1766
7	0.4021	2.0604
8	0.4317	1.9704
9	0.4573	1.8983
10	0.4795	1.8390
11	0.4992	1.7893
12	0.5167	1.7468
13	0.5325	1.7100
14	0.5467	1.6778
15	0.5597	1.6493
16	0.5716	1.6239
17	0.5825	1.6011
18	0.5927	1.5804
19	0.6021	1.5616
20	0.6108	1.5444
21	0.6190	1.5286
22	0.6267	1.5140
23	0.6339	1.5005
24	0.6407	1.4879
25	0.6471	1.4762
26	0.6532	1.4652
27	0.6590	1.4549
28	0.6645	1.4453
29	0.6697	1.4362
30	0.6747	1.4276
31	0.6795	1.4194
32	0.6840	1.4117
33	0.6884	1.4044
34	0.6925	1.3974
35	0.6965	1.3908
36	0.7004	1.3844
37	0.7041	1.3784
38	0.7077	1.3726
39	0.7111	1.3670
40	0.7144	1.3617
41	0.7176	1.3566
42	0.7207	1.3517
43	0.7237	1.3470
44	0.7266	1.3425
45	0.7294	1.3381
46	0.7321	1.3339
47	0.7348	1.3298
48	0.7373	1.3259
49	0.7398	1.3221
50	0.7422	1.3184

number of events	95% Confidence Interval, Lower Limit Factor	95% Confidence Interval, Upper Limit Factor
51	0.7446	1.3148
52	0.7468	1.3114
53	0.7491	1.3080
54	0.7512	1.3048
55	0.7533	1.3016
56	0.7554	1.2986
57	0.7574	1.2956
58	0.7593	1.2927
59	0.7612	1.2899
60	0.7631	1.2872
61	0.7649	1.2845
62	0.7667	1.2820
63	0.7684	1.2794
64	0.7701	1.2770
65	0.7718	1.2746
66	0.7734	1.2722
67	0.7750	1.2700
68	0.7765	1.2677
69	0.7781	1.2656
70	0.7795	1.2634
71	0.7810	1.2614
72	0.7824	1.2593
73	0.7838	1.2573
74	0.7852	1.2554
75	0.7866	1.2535
76	0.7879	1.2516
77	0.7892	1.2498
78	0.7905	1.2480
79	0.7917	1.2463
80	0.7929	1.2446
81	0.7941	1.2429
82	0.7953	1.2413
83	0.7965	1.2397
84	0.7976	1.2381
85	0.7988	1.2365
86	0.7999	1.2350
87	0.8010	1.2335
88	0.8020	1.2320
89	0.8031	1.2306
90	0.8041	1.2292
91	0.8051	1.2278
92	0.8061	1.2264
93	0.8071	1.2251
94	0.8081	1.2237
95	0.8091	1.2224
96	0.8100	1.2212
97	0.8109	1.2199
98	0.8118	1.2187
99	0.8128	1.2175
100	0.8136	1.2163